Accounting for Behavioral Response to Capture when Estimating Population Size from Hair Snare Studies with Missing Data

Ben C. Augustine[a]

Catherine A. Tredick[b]

Simon J. Bonner[c]

[a]Department of Fish and Wildlife Conservation, Virginia Tech, Blacksburg, VA 24061

[b] Environmental Science Program, The Richard Stockton College of New Jersey, Galloway, NJ 08205

[a]Department of Statistics, University of Kentucky, Lexington, KY 40536-0082

Corresponding email: baugusti@vt.edu

December 12, 2014

Summary

1. Hair snares have become an established method for obtaining mark-recapture data for population size estimation of Ursids and have recently been used to study other species including other carnivores, small mammals, and ungulates. However, bias due to a behavioral response to capture in the presence of missing data has only recently been recognized and no statistical methodology exists to accommodate it. In a hair snare mark-recapture experiment, data can be missing if animals encounter a hair snare without leaving a hair sample, poor quality samples are not genotyped, a fraction of all samples collected are genotyped due to cost considerations (subsampling), and/or not all genotyped hair samples provide an individual identification. These are all common features of hair snare mark-recapture experiments.

2. Here, we present methodology that accounts for a behavioral response to capture in the presence of missing data from 1) subsampling and 2) failure of hair samples to produce an individual identification. Four subprocesses are modeled–animal capture, hair deposition, researcher subsampling, and DNA amplification with key parameters estimated from functions of the number of hair samples left by individuals at traps. We assess the properties of this methodology (bias and interval coverage) via simulation and then apply this methodology to a previously published data set.

3. Our methodology removes bias and provides nominal interval coverage of population size for the simulation scenarios considered. In the example data set, we find that removing 75% of the hair samples leads to a 40% lower estimate of population size. Our methodology corrects about half of this bias and we identify a second source of bias that has not previously been reported associated with differential trap visitation rates among individuals within trapping occasions.

4. Our methodology will allow researchers to reliably estimate the size of a closed population in the presence of a behavioral response to capture and missing data for a subset of missing data scenarios. It also provides a framework for understanding this generally unrecognized problem and for further extension to handle other missing data scenarios.

KEYWORDS: Behavioral response, DNA, closed population, hair snare, mark-recapture, missing data

1. INTRODUCTION

Mark-recapture experiments for many taxa increasingly rely on DNA samples for individual identification, with hair samples being one of the main sources of DNA. Identification from DNA in hair samples has been used to study at least 22 species of carnivores (Kendall & McKelvey, 2008) and these methods have recently been applied to small mammals (Henry & Russello, 2011) and ungulates (Belant *et al.*, 2007). For some taxa such as Ursids, DNA identification from hair samples is the dominant method for mark-recapture studies and the resulting population size estimates are used to inform the management of small, extinction-prone populations (e.g. Tredick & Vaughan, 2009; Frary *et al.*, 2011). Given the prevalence of these methods and the importance of reliable population size estimates, potential sources of bias in this methodology need to be understood and where possible, accounted for by extending current models or modifying the experimental design. Several sources of bias have been previously investigated (Roon *et al.*, 2005; Dreher *et al.*, 2009; Laufenberg *et al.*, 2013), but bias due to missing data in the presence of a behavioral response to capture has gone largely unnoticed (but see Laufenberg *et al.*, 2013). This is important because behavioral responses to capture and missing data are both common in hair snare studies.

A behavioral response to capture in hair snare experiments has been documented (e.g. Tredick *et al.*, 2007; Royle *et al.*, 2011) and is likely to occur in many hair snare sampling designs. Because traps are novel, may be associated with humans, and may be uncomfortable to enter (e.g. barbed wire), animals may be reluctant to enter them. To counter this, hair snares are usually baited with either a food reward or a scent (Kendall & McKelvey, 2008) so that animals have an incentive to enter the traps. If food rewards are used, animals may perceive the food reward as worth the discomfort and novelty of entering a hair snare trap and become trap happy. Alternatively, they may not perceive the food reward as worth the trouble and become trap shy. If scents are used, animals may become trap shy after realizing there is no reward associated with the scent (Brian Dreher pers. comm.).

The standard approach to modeling a behavioral response to capture in closed populations is to use model M_b (Otis *et al.*, 1978), which can provide unbiased population size estimates when the magnitude of the behavioral response does not vary among individuals or across time and no missing data are present (the likelihood for M_b can be found in Appendix A). The parameters of M_b are N, the population size, p, the probability of first capture, and c, the probability of subsequent

captures. The number of capture occasions is represented by t and $\boldsymbol{\omega}$ is the matrix of capture histories. Let M_j denote the number of marked individuals in the population on occasion j and m_j denote the number of marked individuals captured on occasion j. The sufficient statistics computed from $\boldsymbol{\omega}$ are $M_\cdot = \sum_{j=1}^{t} M_j$, $m_\cdot = \sum_{j=1}^{t} m_j$ and the total number of individuals captured in the experiment, M_{t+1}.

In this paper, we consider two sources of missing data that may bias estimates of the size of a closed population when there is a behavioral effect. First, a researcher may not genotype all hair samples due to cost considerations. This practice is especially common in studies of black bears (*Ursus americanus*), (Tredick *et al.*, 2007; Settlage *et al.*, 2008), in which hundreds or thousands of hair samples may be collected and only a small proportion can be genotyped. Second, not all genotyped samples will produce an individual identification. Common causes of sample failure are poor quality samples and hair samples containing DNA from more than one individual (Waits & Paetkau, 2005). A third source of missing data that we do not consider is that animals may encounter a hair snare without leaving a hair sample. We also do not address genotyping errors leading to incorrect individual identifications as proper lab protocol can minimize their prevalence to negligible levels, at least in studies using multiple plucked hair samples to obtain DNA (Paetkau, 2003; Roon *et al.*, 2005). We comment further on both of these issues in the Discussion.

Previous studies of the effects of missing data on estimates of population size have focused on three particular sources of bias: 1) interactions between missing data and individual misidentification due to errors in the DNA amplification process (Dreher *et al.*, 2009), 2) reduced number of samples leading to the selection of overly-simple models (Laufenberg *et al.*, 2013), and 3) reduced capture probabilities leading to poor estimator performance (Tredick *et al.*, 2007). However, the effects of missing data in the presence of a behavioral response to capture have not been investigated in detail (Laufenberg *et al.*, 2013) and no methodology exists for obtaining reliable population size point and interval estimates if this occurs.

In Appendix A we describe analytical methods to approximate the bias of $\hat{N}$ from model M_b when missing data are ignored. In general, failure to account for missing data will positively bias $\hat{N}$ if individuals display a trap happy response and negatively bias $\hat{N}$ if individuals display a trap shy response. In addition, the magnitude of the behavioral response ($|p - c|$) is necessarily underestimated when data are missing. To see this, let p_{obs} be the probability that a previously

uncaptured individual is captured and identified and c_{obs} be the probability that an individual is recaptured and identified. If δ is the proportion of data that is not missing due to subsampling or amplification failure then $p_{obs} = \delta p$ and $c_{obs} = \delta c$. If $\hat{p}_{obs}$ and $\hat{c}_{obs}$ are unbiased estimates of p_{obs} and c_{obs} then the estimate of the behavioral effect is $|\hat{p}_{obs} - \hat{c}_{obs}|$ and $E[|\hat{p}_{obs} - \hat{c}_{obs}|] = |\delta p - \delta c| < |p - c|$ unless $\delta = 1$. Note, the analytical methods in Appendix A cannot be used in practice as they depend upon the unknown parameters.

Here, we present methodology that allows researchers to fit M_b in the presence of missing data by explicitly modeling the hair sample deposition, subsampling, and DNA amplification processes. We assess the properties of the methodology under different missing data scenarios via simulation and compare the results to those obtained when naively fitting M_b. Then, we also apply the methodology to data from a previous study of black bears which showed that subsampling decreased $\hat{N}$ to levels up to $\sim 38\%$ below $\hat{N}$ estimated from the complete data set. This study used M_b fit via maximum likelihood which ignores missing data (Tredick *et al.*, 2007). The population under study was estimated to exhibit a trap shy response, which as we show in Appendix A, should produce a negative bias in $\hat{N}$ in the presence of missing data. Tredick *et al.* (2007) mis-attributed the negative bias to poor estimator performance resulting from the low capture probabilities obtained after subsampling. In this paper, we will use this example to demonstrate our methodology. Of particular interest is whether the magnitude of the behavioral response in this experiment can fully explain the observed bias in the estimate of population size, given the amount of missing data.

2. METHODS - MODEL DESCRIPTION

The model we developed can be separated into four processes–animal capture, hair deposition, subsampling, and DNA amplification. We will discuss each in turn. A list of terms and definitions can be found in Table 1 and the full model is depicted in Figure 1. Starting with the animal capture process, we make three assumptions:

(1.1) Population closure

(1.2) Constant capture probability, p, and recapture probability, c across individuals and time periods

(1.3) Capture events are independent

(1.4) Individuals can be captured in at most 1 trap per occasion

Let $\boldsymbol{\omega}$ be the matrix of unobserved capture histories with ω_{ij} being 1 if individual i was captured on occasion j and 0 otherwise. According to the assumptions above, $\omega_{ij} \sim \text{Bern}(q_{ij})$ where $q_{ij} = p$ if an individual i has not been captured before occasion j and c otherwise. Assumption 1.2 is made for convenience and can be relaxed with standard methods for modeling time effects and individual heterogeneity (Otis *et al.*, 1978) or by using individual covariates, if available. The hair deposition process makes three assumptions:

(2.1) Hair samples are left in discrete units, such as all hairs left on one barb, and remain on the barb until the researcher collects the sample

(2.2) Conditional on visiting a trap, animals leave hair samples according to a zero-truncated Poisson distribution.

(2.3) The expected number of hair samples left conditional upon visiting a trap does not depend on the individual, trap, or trapping occasion

Let $\mathbf{S}$ be the matrix containing the unobserved number of hair samples left by each individual on each occasion. According to the assumptions above, conditional on visiting a trap, the number of hair samples left at a trap for individual i captured at time j, S_{ij}, follows a zero-truncated Poisson distribution with parameter λ so that

$$P(S_{ij}|\omega_{ij}) = \begin{cases} \frac{\lambda^{S_{ij}} \exp(-\lambda)}{S_{ij}!(1-\exp(-\lambda))} & \omega_{ij} = 1, S_{ij} > 0 \\ 0 & \omega_{ij} = 1, S_{ij} = 0 \\ 0 & \omega_{ij} = 0, S_{ij} > 0 \\ 1 & \omega_{ij} = 0, S_{ij} = 0 \end{cases} \tag{1}$$

The appropriateness of this distribution can be checked using goodness of fit tests (Best *et al.*, 2007) and other positive count distributions can be used if the zero-truncated Poisson is found to be inappropriate. Note that the zero truncation implies that an individual necessarily leaves at least one hair sample when it enters a trap. This assumption is required to ensure that the model is identifiable. The subsampling process makes one assumption:

(3.1) On each occasion, hair samples from all traps are pooled and a simple random sample is retained with known probability δ

Let $\mathbf{U}$ be the matrix containing the number of hair samples retained for genetic analysis after subsampling for each individual on each occasion. According to the assumptions above, the number of hair samples remaining in the subsample for individual i captured at time j, denoted by U_{ij}, are each distributed as a binomial random variable with size S_{ij} and probability δ. Other subsampling methods are possible and some alternatives are considered in the Discussion. The DNA amplification process makes three assumptions:

(4.1) All samples produce individual identifications with a unknown probability α, which does not vary by hair sample, individual or trapping occasion

(4.2) No false identifications occur (no allelic dropout; Taberlet (1996); or shadow effect; Mills *et al.* (2000))

Let $\mathbf{R}$ be the matrix containing the number of positive identifications. According to the assumptions above, R_{ij}, the number of hair samples from individual i which are in the genotyped subsample at time j are each distributed as a binomial random variable of size U_{ij} with probability α. Under this model, both S_{ij} and U_{ij} are unobserved, while the observed data are R_{ij}, $\boldsymbol{S}_{\boldsymbol{\cdot}} = (S_{\cdot 1}, \ldots, S_{\cdot t})'$, a vector of length t containing the number of hair samples collected on each occasion, and $\boldsymbol{U}_{\boldsymbol{\cdot}} = (U_{\cdot 1}, \ldots, U_{\cdot t})'$, a vector of length t containing the number of hair samples collected on each occasion that are retained in the subsample. Assumption 4.2 is made for convenience and can be relaxed if necessary (e.g. Link *et al.*, 2010).

We fit this model in a Bayesian framework using the complete data likelihood (CDL) and data augmentation (Tanner & Wong, 1987). We used a custom built MCMC sampler in order to enforce constraints imposed by $\boldsymbol{S}_{\boldsymbol{\cdot}}$ and $\boldsymbol{U}_{\boldsymbol{\cdot}}$ (see Appendix B for details). The data augmentation procedure allowed for the estimation of the unknown multinomial index parameter, N, by factoring the multinomial as the product of a binomial modeling the number of individuals in the population and a multinomial modeling the capture history of each individual, both with fixed size (Royle & Dorazio, 2008). The observed capture histories were augmented with $M - M_{t+1}$ "pseudo-individuals" having all zero capture histories, and individuals in the augmented population were included in the true population with probability ψ. Latent indicator variables $z_i \sim \text{Bern}(\psi)$ $(i = 1, 2, ..., M)$ determined

which individuals were in the population and the posterior distribution of N was approximated by calculating $N = \sum_{i=1}^{M} z_i$ on each iteration. We used a Beta(0.001,1) prior for ψ, inducing the scale prior on N which has been shown to avoid unacceptable behavior sometimes encountered with the discrete uniform prior (Link, 2013). Priors for both p and c were Uniform(0,1). Two versions of the model were considered–M_{b2}, which accounts for researcher subsampling, and M_{b3}, which accounts for both researcher subsampling and failure of hair samples to produce an individual identification.

The CDL for the more general model (see Table 1 for notation review), M_{b3}, is

$$L_{M_{b3}}(\boldsymbol{z}, \boldsymbol{\omega}, \boldsymbol{S}, \boldsymbol{U}, p, c, \lambda, \alpha | \boldsymbol{R}, \boldsymbol{S.}, \boldsymbol{U.}) = P(\boldsymbol{S.}, \boldsymbol{U.} | \boldsymbol{S}, \boldsymbol{U}) P(\boldsymbol{R}, \boldsymbol{U} | \boldsymbol{S}, \alpha) P(\boldsymbol{z}, \boldsymbol{\omega}, \boldsymbol{S} | \psi, p, c, \lambda)$$

where:

$$P(\boldsymbol{z}, \boldsymbol{\omega}, \boldsymbol{S} | \psi, p, c, \lambda) = \prod_{i=1}^{M} \psi^{z_i} (1-\psi)^{z_i} \cdot \prod_{i=1}^{M} \prod_{j=1}^{t} q_{ij}^{\omega_{ij}} (1-q_{ij})^{1-\omega_{ij}} P(S_{ij} | \omega_{ij})$$

models capture and sample deposition,

$$P(\boldsymbol{R}, \boldsymbol{U} | \boldsymbol{S}, \alpha) = \prod_{i=1}^{M} \prod_{j=1}^{t} \delta^{U_{ij}} (1-\delta)^{S_{ij}-U_{ij}} \alpha^{R_{ij}} (1-\alpha)^{R_{ij}}$$

models the processes of subsampling and genotyping failure, and

$$P(\boldsymbol{S.}, \boldsymbol{U.} | \boldsymbol{S}, \boldsymbol{U}) = \prod_{j=1}^{t} I\left(\sum_{i=1}^{M} S_{ij} = S_{\cdot j} \text{ and } \sum_{i=1}^{M} U_{ij} = U_{\cdot j} \right)$$

ensures that the number of samples deposited and subsamples genotyped on each occasion match the observed values. Here $I(\cdot)$ is the indicator function. The CDL for the reduced model, M_{b2}, that accounts for missing data from the subsampling process only is the same as above after setting $\alpha = 1$.

3. SIMULATION STUDY

Simulations of closed populations of size 250 were conducted to assess the frequentist properties of the methodology (bias and interval coverage) and to compare the performance to naively fitting

M_b. In Simulation 1 we considered 18 scenarios in which data were missing only due to systematic subsampling by researchers so that M_{b2} was the correct model. Data were simulated from M_{b2} with different values of p, c, δ, and λ and then both M_b and M_{b2} were fit to the data (see Table 2 for specific parameter combinations). Of these 18 scenarios, 9 considered a trap-happy response and 9 considered a trap-shy response. The magnitude of behavioral response ($|p - c|$) was either 0.2 or 0.4 and 6 capture occasions were simulated.

In Simulation 2 we considered both subsampling and failure of hair samples to produce an individual identification so that M_{b3} was the correct model. We chose 6 scenarios to produce the same level of missing data as the most extreme subsampling-only scenarios ($\delta = 0.5$, $\lambda = 1$), but with half of the missing data due to subsampling and half to genotyping failure, achieved by setting $\delta = \sqrt{0.5}$ and $\alpha = \sqrt{0.5}$. In these scenarios, data were simulated from M_{b3} and M_{b3} was fit to the data. We did not fit M_b in these scenarios because this replicates the results of the previous simulation. If the data are randomly subsampled by two binomial processes, then the overall missing data process is still binomial with $p = \delta\alpha = (\sqrt{0.5})(\sqrt{0.5}) = 0.5$ as in Simulation 1. For all simulations, each scenario was repeated 100 times and the following summary statistics were calculated for N: mean posterior mode, 95% highest posterior density (HPD) credible interval coverage of the true parameter, mean 95% HPD credible interval width, and mean estimated behavioral response.

3.1 Simulation 1

Naively fitting M_b in the presence of missing data and a behavioral response produced positively-biased estimates of N in trap-happy scenarios and negatively-biased estimates of N in trap-shy scenarios (Table 2). In trap-happy scenarios, bias ranged from +1% to +18% and in trap-shy scenarios, bias ranged from -1% to -13%. For both trap response types, bias was greater when capture and recapture probabilities were lower, when the behavioral response was larger, and when λ was smaller. As the level of missing data increased or as λ decreased, credible interval coverage decreased and credible interval width increased. The increase in credible interval width was greater in trap-happy scenarios, leading to a smaller reduction in credible interval coverage than in the trap-shy scenarios.

Model M_{b2} substantially reduced bias: posterior modes for N were essentially unbiased for both

trap-happy and trap-shy scenarios. Coverage of the 95% HPDs was close to nominal for both types of trap response with the mean coverage probability across the 6 trap-happy scenarios being 0.957 and across the trap-shy scenarios being 0.965 (compared to 0.907 and 0.320, respectively for M_b). In the trap-shy scenarios credible interval widths were wider than those for M_b on average. These differences increased as the level of missing data increased and as λ decreased.

Estimates of the behavioral response were also negatively biased when naively fitting M_b. Bias was greater when capture and recapture probabilities were higher, when the behavioral response was larger, and when λ was smaller. Bias in missing data scenarios ranged from -16% to -41%. M_{b2} effectively removed bias with a mean bias across all scenarios of +0.4%.

3.2 Simulation 2

In Simulation 2, M_{b3} estimates of N were effectively unbiased with near nominal credible interval coverage (mean of 0.96 across all 6 scenarios see Table 2 for full results). As before, these differences increased as the level of missing data increased and as λ decreased. Additionally, these differences were larger for trap-happy scenarios. Model M_{b3} largely removed bias in the behavioral response with a mean bias of +2%.

4. EXAMPLE

We applied our methodology to a data set from a closed population black bear hair snare study conducted on the Pocosin Lakes National Wildlife Refuge in Northeastern North Carolina (Tredick *et al.*, 2007) that was not originally subsampled. Details relevant to the current study will be provided here–see Tredick *et al.* (2007) for a complete description of the study. Thirty-three baited barbed-wire hair snare traps were checked over 8 capture occasions, yielding 85 unique individual identifications. Of the 468 hair samples collected, 85% provided an individual identification. The data were originally analyzed using CAPTURE (White, 1982) and evidence was found for individual heterogeneity in capture probabilities, time effects, and a trap-shy behavioral response.

Using the subsampling method assumed by our model, we simulated the subsampling process at four levels ($\delta = 1, 0.75, 0.50$ and 0.25). Note, this approach is slightly different than used in Tredick *et al.* (2007). We simulated subsampling before DNA amplification since researchers cannot know ahead of time which samples will produce an individual identification while Tredick *et al.* (2007)

simulated subsampling after DNA amplification. Before subsampling, hair samples from individuals that were captured at multiple traps on the same occasion were combined by individual. At each stage, we fit M_b and M_{b3} and recorded the posterior mode and 95% HPD interval for N. The entire process was repeated 100 times to accommodate variability in the subsampling process. Time effects and individual heterogeneity were not considered. Therefore, our population size estimates for the true population will be biased, but we are only interested in how estimates change with increasing levels of missing data and these effects should not introduce bias as the level of missing data increases. Coverage and relative bias were calculated using the "best estimate" of N for this data, which was the estimate from M_b when $\delta = 1$. We believed using the best estimate to calculate bias and coverage will give reasonably accurate results since the 95% credible interval for N when no data are missing is narrow (83-89).

As in the previous simulations, naively fitting M_b in the presence of missing data and a trap-shy behavioral response to capture produced negatively-biased estimates of N (Table 3). Relative bias from the best estimate increased from 7% to 40% as data were progressively subsampled and credible interval coverage of the best estimate was reduced to 0.02 when δ=0.25. Model M_{b3} performed substantially better than M_b, removing about half of the bias and increasing credible interval coverage of the best estimate to 0.47 when δ=0.25. We explore possible reasons we could not remove the majority of bias in Appendix C.

5. DISCUSSION

We have demonstrated analytically and through simulation that M_b produces biased estimates of population size in the presence of missing data, showed how this bias is introduced, and provided methodology to correct this bias in the presence of two sources of missing data and under one model of subsampling. We also demonstrated that M_b underestimates the magnitude of the behavioral response in the presence of missing data, making it less likely that a behavioral response will be detected. We showed that our methodology provides essentially unbiased estimates of both N and the behavioral response and near nominal frequentist interval coverage probabilities for N in the range of sampling scenarios we considered when the model assumptions are satisfied. We demonstrated that about 50% of the total negative bias observed in Tredick *et al.* (2007) can be explained by a trap-shy behavioral response in the presence of missing data. We also

identified a second source of bias that occurs in the presence of missing data – a specific form of individual heterogeneity in capture probability (see Appendix C). Individual heterogeneity in capture probability itself is not problematic in the presence of missing data (confirmed by simulation results not presented here); however, if individuals with higher capture probabilities leave more hair samples per occasion than those with lower capture probabilities, the latter will drop out of the observed sample faster than the former, resulting in less observed heterogeneity and a mean capture probability that is biased high. As a result, $\hat{N}$ will be biased low even if individual heterogeneity is modeled. This source of bias appeared to explain another 17.5 % of the total negative bias in the Tredick *et al.* (2007) data. We suspect this pattern is caused by bears with higher capture probabilities visiting more traps per occasion which has been documented elsewhere (e.g. Van Manen *et al.*, 2012).

We were unable to account for about 32.5% of the total negative bias relative to the best estimate due to missing data in the Tredick *et al.* (2007) data set. The fact that the magnitude of the behavioral response was not underestimated by M_b when missing data were present and the behavioral response did not remain constant as the level of missing data were increased using M_{b3} suggests that M_b does not closely approximate the data generating process. It may be that subsampling is interacting with other sources of bias in this data set or even that the original observed behavioral response is largely explained by another source of bias. For example, closure may have been violated and missing data may be interacting with Markovian movement on and off the grid or with permanent emigration/immigration since this experiment was started during the time of year subadult males are dispersing (see Kendall, 1999). Due to this uncertainty, both our estimate and the original should be treated cautiously.

Our methodology was successful under the assumptions made regarding the hair sample subsampling and DNA amplification processes, and further extensions can make this methodology more widely applicable. First, our methodology could be extended to accommodate the correlation between individual capture probabilities and the number of hair samples left upon capture using the Poisson encounter model of Royle *et al.* (2009) to model both the distributions of the number of captures per occasion and the number of hair samples left conditional upon capture. This could also address overdispersion in S_{ij} due to pooling across traps if the number of hair samples left at individual traps are well modeled by a Poisson. Second, we have modeled subsampling as

a simple random sample, but subsampling is often conducted in other ways (e.g. Tredick *et al.*, 2007; Settlage *et al.*, 2008; Dreher *et al.*, 2009). Researchers frequently subsample in a manner that maximizes the probability of identifying unique individuals. Since samples found at the same trap/occasion are more likely to be from the same individual, genotyping multiple samples from the same trap/occasion leads to diminishing returns in precision and accuracy (Dreher *et al.*, 2009). Therefore, researchers frequently take a systematic sample, for example, one sample from each trap/occasion or one sample from a subset of traps on each occasion (e.g. Settlage *et al.*, 2008). The strategy of taking a fixed number of samples from each trap or a subsample of traps can be accommodated by implementing different versions of the subsampling model.

Third, the subsampling process often contains a nonrandom component. Since hair samples vary in their probability of producing an individual identification (David Paetkau pers. comm.), researchers often send only high quality samples to the lab (Tredick *et al.*, 2007; Wegan *et al.*, 2012). The quality of samples varies by the number of hairs with roots per sample, the type and duration of environmental conditions samples were exposed to before collection (David Paetkau pers. comm.), and the time of year the samples were collected (Wegan *et al.*, 2012). If all hair samples had an equal probability of being in the subsample regardless of sample quality, amplification rates could be modeled as a function of sample quality covariates. Alternatively, our methodology can remove bias by only modeling the high quality samples, but bias will remain to the extent that low quality samples were left upon first capture and to the extent that there still remains variability in α among the high quality samples.

Finally, the DNA amplification process can be more complex than we assumed. We ignored the occurrence of genotyping error, specifically, individuals in the population having the same genotype (Shadow effect; Mills *et al.*, 2000) and identification of false individuals due to allelic dropout and false amplification (Taberlet, 1996). Roon *et al.* (2005) demonstrated via simulation that with appropriate error-checking protocols, bias from these errors can be minimized at error rates typical of studies using multiple plucked hairs to obtain DNA samples (e.g. Paetkau, 2003). If bias from these errors is thought to be large enough to warrant correction, existing models to correct this bias (e.g. Link *et al.*, 2010) could be extended to handle missing data in the presence of a behavioral response to capture.

We also did not investigate missing data due to individuals undergoing a behavioral response

without leaving a hair sample because there is no information available in the typical hair snare mark-recapture study to model this source of bias. Our lack of knowledge of the magnitude of this source of missing data in typical experiments leads to substantial uncertainty about how biased experiments with behavioral responses and data subsampling may be. We have found only one attempt to estimate this quantity (Boulanger *et al.*, 2004). Using their top model, Boulanger *et al.* (2004) estimated the probability of leaving at least one hair sample and at least one of those hair samples producing an individual identification conditional on visiting a trap was 0.49 (C.I. = 0.26-0.72). The estimated success rate for hair samples producing an individual identification, $\hat{\alpha}$, was not estimated and we do not know how many hair samples bears left when they left ≥ 1 sample. Using the values of $\hat{\alpha}$ (0.85) and $\hat{\lambda}$ (1.8) observed in the data from Tredick *et al.* (2007) and assuming our model structure, about 6% of bears leaving at least one hair sample will not produce an individual identification, so we can estimate the probability of leaving at least one hair sample conditional upon visiting a trap at 0.55. If this estimate is accurate and these bears underwent a behavioral response, this source of missing data could introduce substantial bias, indicating that bias may be of concern even if no subsampling takes place. In bear hair snare studies, this problem may be reduced by using two strands of barbed wire rather than a single strand (Boulanger *et al.*, 2006) and similar strategies for making traps more effective may exist for other species. The additional samples from more effective traps can reduce bias by making the missing data explicit, allowing it to be modeled so the only added cost would be more expensive traps and in some cases, installation time.

While we have focused on trying to reduce bias by modeling the behavioral response, hair deposition, and DNA amplification processes, another strategy for reducing bias is to reduce the magnitude of the behavioral response or try to remove it completely. Moving trap locations between occasions has been successful in reducing individual heterogeneity in capture probabilities and arguably reducing the magnitude of negative behavioral response to capture (Boulanger *et al.*, 2006). We think that missing data bias due to both behavioral response to capture and a correlation between individual capture probability and number of hair samples deposited makes the case for this sampling strategy even more compelling.

As argued in the Introduction, a behavioral response to capture should be expected in hair snare experiments, but of most importance is the magnitude of the effect. As we demonstrated, M_b

underestimates this magnitude in the presence of missing data. Further, rarefied data leads to the selection of simpler models (Laufenberg *et al.*, 2013). Together, the analyst is left with less power to detect a behavioral response and if detected, the magnitude will be underestimated. Therefore, the prevalence of support for M_b and the magnitudes of behavioral responses observed in the literature are unlikely to be reliable indicators of how frequent and large behavioral responses are in typical hair snare experiments. In order to assess the prevalence and magnitude of behavioral responses in typical hair snare studies, methods that model the missing data need to be widely applied.

On a final note, we considered the dominant model for behavioral responses in classical mark-recapture methodology for closed populations using hair snares for individual identification. However, the mechanisms of bias we identified should apply to other behavioral response models (e.g. Yang & Chao, 2005; Hwang & Huggins, 2011; Ramsey & Severns, 2010), behavioral responses in spatial mark recapture models (e.g. Royle *et al.*, 2011), and in camera trap studies to the extent there are missing data (e.g. photographs that do not produce an individual identification) and a behavioral response to capture. It may be worthwhile to investigate the importance of missing data in these other contexts.

6. DATA ACCESSIBILITY

The R code used to simulate from and fit M_{b2} and M_{b3} and the example data set are available in online supporting information.

7. TABLES

Table 1: Model notation

Term	Definition
M	Size of the super population
N	Size of the population
ψ	Probability that an individual in the super population is included in the population
z_i	1 if individuals in superpopulation are in the population, 0 otherwise
p	Probability of first capture
c	Probability of subsequent capture
q_{ij}	Probability of capture for individual i on occasion j. Each element is either p or c, depending on capture history before j
ω_{ij}	1 if individual i was captured on occasion j, 0 otherwise
λ	Parameter determining the distribution of hair samples left conditional on an individual encountering a trap
δ	Sample retention probability during subsampling
α	Probability a sample will produce an individual identification given that it is genotyped
S_{ij}	Number of hair samples collected from individual i on occasion j
U_{ij}	Number of hair samples collected from individual i on occasion j that remain after subsampling
R_{ij}	Number of hair samples collected from individual i on occasion j that remain after subsampling and produce an individual identification
$S_{\cdot j}$	Number of hair samples collected from all individuals on occasion j
$U_{\cdot j}$	Number of hair samples collected from all individuals on occasion j that remain after subsampling

Table 2: Bias in population size estimates, 95% CI coverage, mean 95% CI width, and bias in mean estimated behavioral response when fitting M_b and M_{b2} to data simulated from M_{b2} and M_{b3} to data simulated from M_{b3}. N=250 for all simulations. $\alpha = \sqrt{0.5}$ for all M_{b3} scenarios.

	Generating Model					Fitting with M_b				Fitting with Generating Model			
						$\hat{N}$			$\lvert\hat{p}$ - $\hat{c}\rvert$	$\hat{N}$			$\lvert\hat{p}$ - $\hat{c}\rvert$
Scenario		p	c-p	λ	δ	% Bias	CI Cov.	CI Width	% Bias	% Bias	CI Cov.	CI Width	% Bias
1	M_{b2}	0.3	+0.2	1	0.5	+14	0.93	228.73	-27	0	0.99	74.40	+1
2	M_{b2}	0.3	+0.2	3	0.5	+4	0.92	80.55	-16	-1	0.94	54.61	-2
3	M_{b2}	0.3	+0.2	3	1.0	0	0.94	43.97	0	0	0.96	42.97	-1
4	M_{b2}	0.5	+0.2	1	0.5	+4	0.94	39.44	-39	0	0.96	20.73	+1
5	M_{b2}	0.5	+0.2	3	0.5	+1	0.95	17.49	-23	0	0.98	12.78	-2
6	M_{b2}	0.5	+0.2	3	1.0	0	0.94	8.35	-4	0	0.94	8.34	+1
7	M_{b2}	0.3	+0.4	1	0.5	+18	0.80	215.22	-32	0	0.92	69.18	+1
8	M_{b2}	0.3	+0.4	3	0.5	+6	0.90	82.77	-17	0	0.95	55.55	+1
9	M_{b2}	0.3	+0.4	3	1.0	-1	0.92	43.96	-8	0	0.93	43.42	0
10	M_{b2}	0.5	-0.2	1	0.5	-10	0.27	32.87	-29	0	1.00	36.67	-1
11	M_{b2}	0.5	-0.2	3	0.5	-4	0.63	17.60	-18	0	0.96	19.98	0
12	M_{b2}	0.5	-0.2	3	1.0	0	0.98	8.39	-1	0	0.95	9.00	-1
13	M_{b2}	0.7	-0.2	1	0.5	-4	0.41	15.17	-41	0	0.97	18.66	+3
14	M_{b2}	0.7	-0.2	3	0.5	-2	0.59	5.02	-25	0	0.97	8.25	+2
15	M_{b2}	0.7	-0.2	3	1.0	0	0.99	1.11	-1	0	1.00	1.11	+3
16	M_{b2}	0.7	-0.4	1	0.5	-13	0.01	12.79	-36	0	0.95	44.84	0
17	M_{b2}	0.7	-0.4	3	0.5	-5	0.01	4.66	-20	0	0.94	12.55	+1
18	M_{b2}	0.7	-0.4	3	1.0	0	0.96	1.10	+1	0	1.00	1.21	-1
1b	M_{b3}	0.3	+0.2	1	$\sqrt{0.5}$					-1	0.97	70.21	+3
4b	M_{b3}	0.5	+0.2	1	$\sqrt{0.5}$					0	0.98	20.56	+1
7b	M_{b3}	0.3	+0.4	1	$\sqrt{0.5}$	See corresponding results above.				-2	0.96	63.06	+3
10b	M_{b3}	0.5	-0.2	1	$\sqrt{0.5}$					-1	0.95	35.88	+3
13b	M_{b3}	0.7	-0.2	1	$\sqrt{0.5}$					0	0.95	18.52	+1
16b	M_{b3}	0.7	-0.4	1	$\sqrt{0.5}$					-1	0.95	35.76	+2

Table 3: Population size estimates, bias (relative to the best estimate of 86), 95% CI coverage, mean 95% CI width, and mean behavioral response estimate when fitting M_b and M_{b3} to the example data set with different levels of missing data. In two scenarios, the empirical distribution of the number of hair samples, R, is replaced to remove additional sources of bias.

	M_b					M_{b3}				
δ	Mean $\hat{N}$	% Bias	CI Cov.	CI Width	Mean $\lvert\hat{p}-\hat{c}\rvert$	Mean $\hat{N}$	% Bias	CI Cov.	CI Width	Mean $\lvert\hat{p}-\hat{c}\rvert$
1.00	86	.	.	5.94	0.13	85	-1	.	2.00	0.17
0.75	80	-7	0.54	6.68	0.15	83	-3	0.88	8.93	0.22
0.50	71	-17	0.07	8.38	0.16	78	-9	0.52	14.05	0.28
0.25	52	-40	0.02	13.81	0.16	69	-20	0.47	25.52	0.31

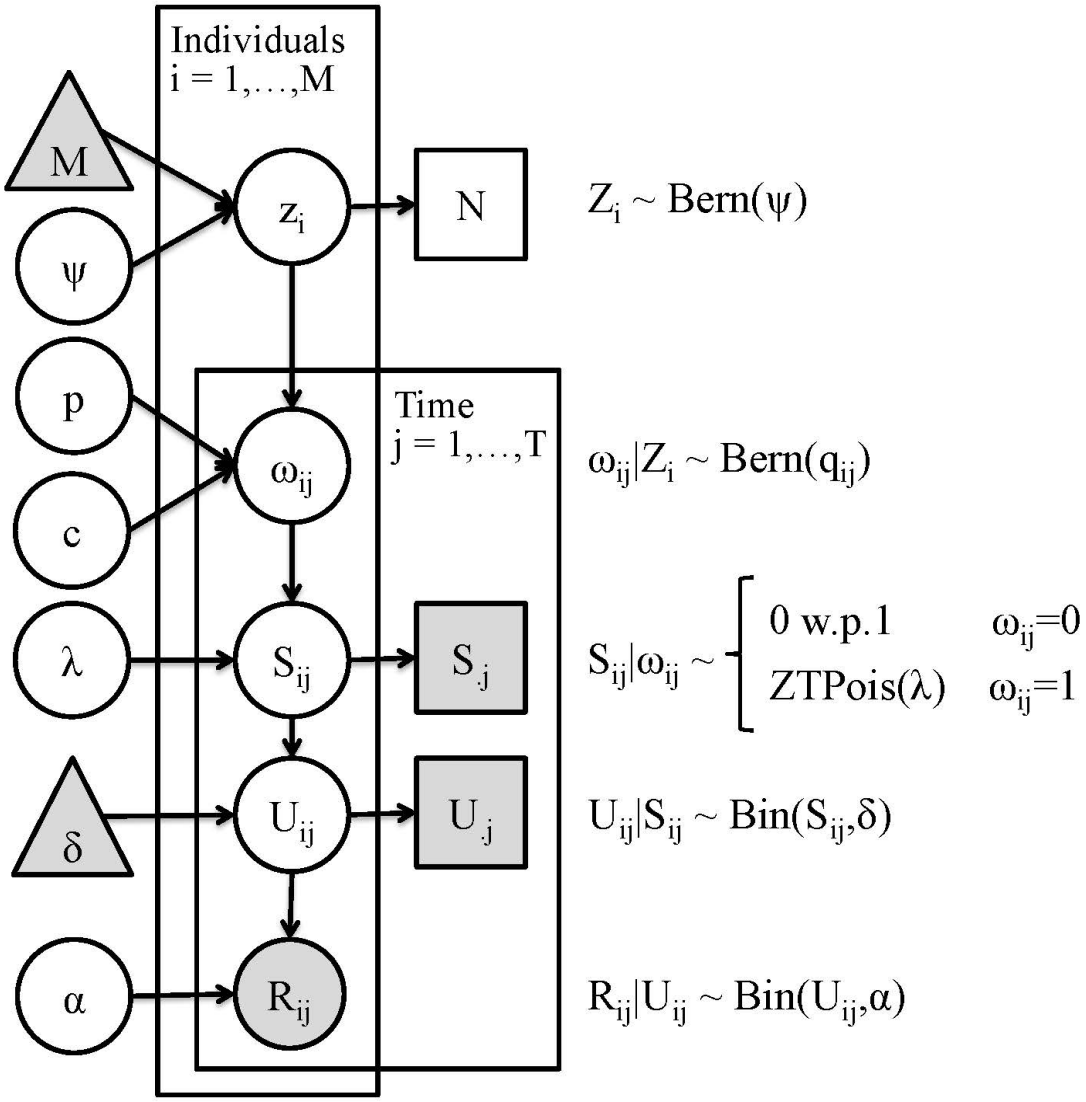

Figure 1: Directed Acyclic Graph of model M_{b3}. Stochastic nodes are represented by circles, fixed parameters are represented by triangles, and deterministic functions of stochastic nodes are represented by squares. Shaded nodes are observed and unshaded nodes are unobserved.

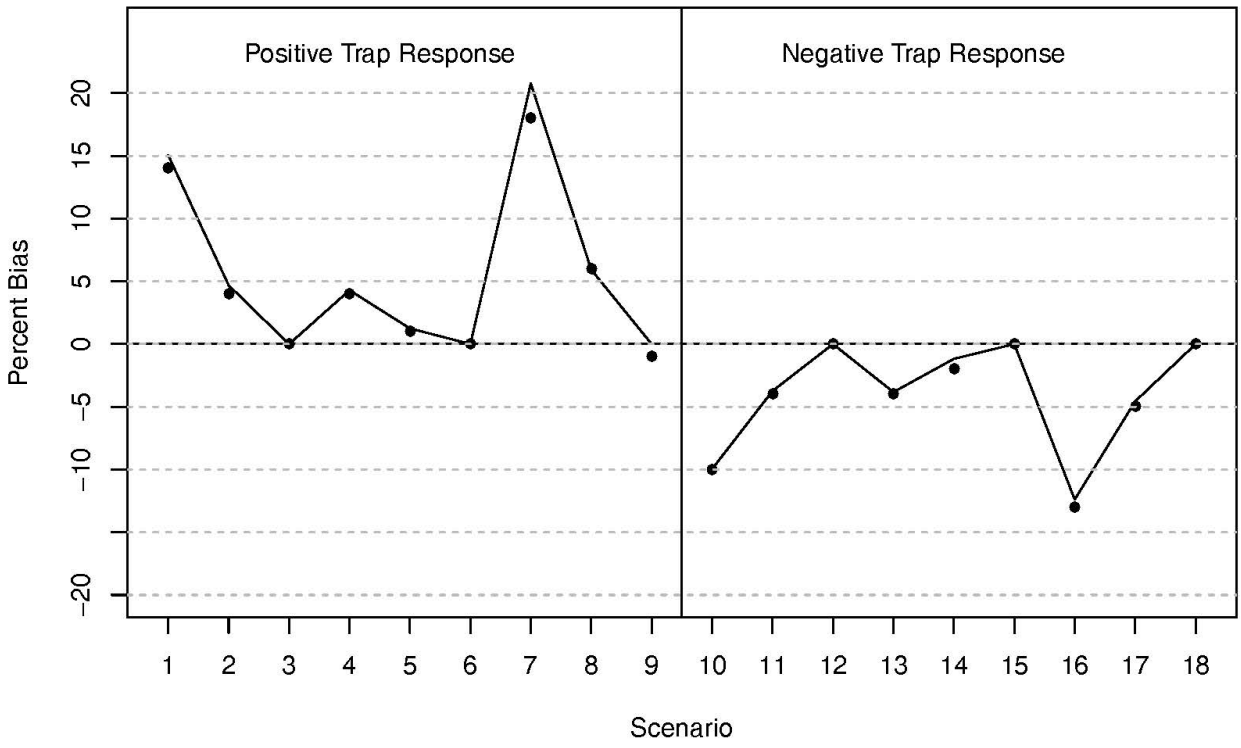

Figure 2: Comparison of the simulated and approximated bias of $\hat{N}$ computed via model M_b in Simulation 1. The points represent the simulated bias for each scenario and the lines represent the approximated bias computed with the method described in Appendix A.1. Scenario numbers correspond to the numbering in Table 2.

APPENDIX A. APPROXIMATE BIAS OF M_b

Here we develop methods to approximate the bias of the estimators of p and N from model M_b to understand how missing data affects these estimators. The likelihood for M_b is

$$L\left[N, p, c|\boldsymbol{\omega}\right] = \frac{N!}{(N - M_{t+1})!}\left(p^{M_{t+1}}(1-p)^{tN-M_{t+1}-M_\cdot}\right)\left(c^{m_\cdot}(1-c)^{M_\cdot - m_\cdot}\right) \tag{A.1}$$

To study the bias of $\hat{N}$ and $\hat{p}$, we consider a two-stage process first maximizing the conditional likelihood for the marked individuals:

$$L(p|M_{t+1}, M_\cdot) = \frac{p^{M_{t+1}}(1-p)^{tM_{t+1}-M_{t+1}-M_\cdot}}{(1-(1-p)^t)^{M_{t+1}}} \tag{A.2}$$

to obtain $\hat{p}$ and then computing the Horvitz-Thompson estimator:

$$\hat{N} = \frac{M_{t+1}}{1-(1-\hat{p})^t}.$$

The results of **?** show that these estimators are asymptotically equivalent to the full maximum likelihood estimators obtained from equation (A.1). However, the asymptotic properties of these estimators are easier to study because the conditional likelihood can be written as the product of a fixed number, M_{t+1}, of independent likelihood contributions.

Let A_i denote the occasion when individual i is first identified (i.e., captured and genotyped) with $A_i = 0$ if the individual is never identified. Then $M_\cdot = \sum_{i=1}^{M_{t+1}}(t - A_i) = M_{t+1}(t - \bar{A})$ where $\bar{A} = \frac{1}{M_{t+1}}\sum_{i=1}^{M_{t+1}} A_i$ represents the mean occasion of first capture for the marked individuals. Substituting into equation (A.2), the conditional likelihood is maximized by equating $\bar{A}$ with its expected value under M_b:

$$\bar{A} = E_{M_b}(A|A \neq 0) = \frac{\sum_{a=1}^{T} a(1-p)^{(a-1)}}{\sum_{j=1}^{T}(1-p)^{(j-1)}}$$

and solving for p. Following **?**, pg. 55, the asymptotic bias of the conditional MLE can be computed as $\text{Bias}(\hat{p}) = p_0 - p$ where p_0 solves the equation:

$$E_{True}(A|A \neq 0) = \frac{\sum_{a=1}^{T} a(1-p_0)^{(a-1)}}{\sum_{j=1}^{T}(1-p_0)^{(j-1)}} \tag{A.3}$$

and

$$E_{True}(A|A \neq 0) = \frac{\sum_{a=1}^{T} aP_{True}(A = a)}{P_{True}(A \neq 0)}$$

represents the expected occasion of first capture under the true data generating model conditional on an individual being captured at least one time. The percent bias in $\hat{p}$ can then be approximated as $\%\text{Bias}(\hat{p}) \approx 100(p_0/p - 1)$.

To study the bias of $\hat{N}$ we consider that $M_{t+1} = \sum_{i=1}^{N} I(A_i \neq 0)$ so that:

$$\hat{N} = \frac{\sum_{i=1}^{N} I(A_i \neq 0)}{1 - (1 - \hat{p})^t}.$$

Application of the weak law of large numbers implies that:

$$\frac{1}{N} \sum_{i=1}^{N} I(A_i \neq 0) \xrightarrow{P} P_{True}(A_i \neq 0)$$

as $N \to \infty$. Combined with the previous result, this provides a formula for the approximate percent bias of $\hat{N}$:

$$\%\text{Bias}(\hat{N}) \approx 100 \left(\frac{P_{True}(A_i \neq 0)}{1 - (1 - p_0)^t} - 1 \right).$$

Note that if the data are truly generated from M_b then $p_0 = p$ and $P_{True}(A_i \neq 0) = 1 - (1 - p)^t$ so that both $\%\text{Bias}(\hat{p}) = 0$ and $\%\text{Bias}(\hat{N}) = 0$.

A.1 Subsampling

Suppose that data is generated from model M_b, with one sample collected per capture, except that samples are selected for genotyping with probability δ. Under this model:

$$P_{True}(A = a) = \begin{cases} p\delta & a = 1 \\ p\delta \left[(1-p)^{(a-1)} + c(1-\delta) \sum_{j=1}^{a-1} (1-p)^{(a-j-1)} (1-c\delta)^{j-1} \right] & a = 2, \ldots, t \end{cases}$$

and $P_{True}(A = 0) = 1 - \sum_{a=1}^{t} P_{True}(A = t)$. For general t, $\%\text{Bias}(\hat{p}) = 0$ and $\%\text{Bias}(\hat{N}) = 0$ can be computed numerically by computing $E_{True}(A|A \neq 0)$ and solving for p_0 in equation (A.3). For

$t = 2$ the equations can be solved explicitly yielding:

$$\%\text{Bias}(\hat{p}) = -100\frac{(1-\delta)c}{p} \quad \text{and} \quad \%\text{Bias}(\hat{N}) = 100\frac{(1-\delta)(c-p)}{p-(1-\delta)c}.$$

Note that $\hat{p}$ has negative asymptotic bias for all values of p, c, and δ in $(0, 1)$, and the magnitude of the bias increases as c/p increases and δ decreases. The asymptotic bias of $\hat{N}$ depends on the behavioral response. A trap happy response, $c > p$, will produce a positive bias while a trap shy response, $c < p$, will produce a negative bias.

The formula for $\%\text{Bias}(\hat{N})$ with $t = 2$ suggests that the direction of the bias might be switched for both trap happy and trap shy responses if $(1-\delta)c > p$. Note however that this only occurs when $\%\text{Bias}(\hat{p}) < -100\%$ so that $E(\hat{p}) < 0$. The problem is essentially that too much subsampling occurs in these cases for M_b to be at all plausible. Under model M_b, $P_{M_b}(A = 1) > P_{M_b}(A = 2)$ for all values of p and c. However, if $\delta < 1 - p/c$ then it is possible that $P_{True}(A = 1) < P_{True}(A = 2)$ under the data generating model with subsampling. In this case, the maximum of the conditional likelihood is not an interior point of the support of p and the conditions for studying the asymptotics of maximum likelihood estimators are broken. In practice, one would restrict $\hat{p}$ to be positive by setting $\hat{p} = \epsilon > 0$ if the conditional likelihood function is monotone decreasing over $(0, 1)$. Given this restriction, $\%\text{Bias}(\hat{p}) > -100$ and the direction of the bias is completely determined by the type of behavioral effect.

A.2 Multiple Samples and DNA Amplification Failure

If animals leave multiple samples on each occasion and/or not all DNA samples produce an individual identification, then the asymptotic bias of $\hat{p}$ and $\hat{N}$ from model M_b can be computed by following the same argument except that δ is replaced by the probability that an individual is identified when captured. Assuming that the number of samples left on capture follows the zero-truncated Poisson distribution with parameter λ, as defined in equation (1), and that errors occur independently with probability α, the probability that a captured individual is identified is:

$$P_{True}(R_{ij} \geq 1 | \omega_{ij} = 1) = \frac{e^{\lambda}(1 - e^{-\lambda\alpha\delta})}{e^{\lambda} - 1}.$$

As before, the approximate bias of $\hat{N}$ and $\hat{p}$ can be computed numerically by computing the expected value of A and solving for N_0 and p_0. Figure 2 compares the approximate bias computed by this method with the empirical bias of the estimates for model M_b obtained from Simulation 1. The results show strong correspondence with only slight deviations at the highest levels of subsampling.

APPENDIX B. MCMC ALGORITHM

Values from the joint posterior of the complete data $\boldsymbol{z}$, $\boldsymbol{S}$, $\boldsymbol{U}$, $\boldsymbol{R}$ and the model parameters ψ, p, c, λ, α are generated via MCMC using the following steps to update these values:

1. Update $\boldsymbol{z}$:

   Gibbs sampling step for each $i = M_{t+1} + 1, \ldots, M$. If $\sum_{j=1}^{t} S_{ij} > 0$ then $z_i = 1$ with probability 1. Otherwise:

   $$z_i \sim \text{Bernoulli}\left(\frac{\psi(1-p)^t}{\psi(1-p)^t + (1-\psi)}\right)$$

2. Update $\boldsymbol{S}$:

   Separate Metropolis-Hastings steps for each occasion. The proposal for occasion j is constructed by reassigning the ungenotyped samples for a randomly selected set of K individuals as follows:

   Given $\boldsymbol{z}$, $\boldsymbol{U}$, and $\boldsymbol{S}^{\text{curr}}$:

   (a) Sample $i_1, \ldots, i_K$ from $\{i : z_i = 1\}$ without replacement.

   (b) Compute $d_k = S_{i_k j} - U_{i_k j}$, $k = 1, \ldots, K$.

   (c) Generate $d'_1, \ldots, d'_K \sim \text{Multinomial}\left(\sum_{k=1}^{K} d_k, (1/K, \ldots, 1/K)\right)$.

   (d) Set $S'_{i_k j} = U_{i_k j} + d'_{i_k}$ $k = 1, \ldots, K$ and $S'_{ij} = S_{ij}^{\text{curr}}$ otherwise.

3. Update $\boldsymbol{U}$: (Model M_{b3} only)

   Separate Metropolis-Hastings steps for each occasion. The proposal for occasion j is constructed by reassigning genotyping failures for a randomly selected set of K individuals as follows:

   Given $\boldsymbol{S}$, $\boldsymbol{R}$, and $\boldsymbol{U}^{\text{curr}}$:

(a) Sample $i_1, \ldots, i_K$ from $\{i : \sum_{j=1}^{t} S_{ij} \geq 1\}$ without replacement.

(b) Compute $d_k = U_{i_k j} - R_{i_k j}$, $k = 1, \ldots, K$.

(c) Generate $d'_1, \ldots, d'_K \sim \text{Multinomial}\left(\sum_{k=1}^{K} d_k, (1/K, \ldots, 1/K)\right)$.

(d) Set $U'_{i_k j} = R_{i_k j} + d'_{i_k}$ $k = 1, \ldots, K$ and $U'_{ij} = S_{ij}^{\text{curr}}$ otherwise.

4. Update ψ, p, c, λ:

   The parameters ψ, p, c, and λ are updated in a joint Metropolis-Hastings step with proposal generated from a multivariate normal distribution centered on the current values:

$$(\text{logit}(\psi', p', c'), \log(\lambda'))' \sim N\left((\text{logit}(\psi^{\text{curr}}, p^{\text{curr}}, c^{\text{curr}}), \log(\lambda^{\text{curr}}))', \Sigma\right).$$

   The variance-covariance matrix of the proposal distribution is tuned so that:

$$\Sigma \approx \frac{2.38^2}{4} \text{Var}\left(\text{logit}(\psi^{\text{curr}}), \text{logit}(p^{\text{curr}}), \text{logit}(c^{\text{curr}}), \log(\lambda^{\text{curr}}) | \boldsymbol{S}\right)$$

   to give the optimal acceptance probability (**??**).

5. Update α: (Model M_{b3} only)

   Gibbs sampling step:

$$\alpha | \boldsymbol{R}, \boldsymbol{U} \sim \text{Beta}\left(1 + \sum_{i=1}^{M}\sum_{j=1}^{t} R_{ij}, 1 + \sum_{i=1}^{M}\sum_{j=1}^{t}(U_{ij} - R_{ij})\right)$$

APPENDIX C.

We identified two potential reasons our methodology did not perform as well in the example as it did in the simulations: first, overdispersion in S_{ij} , possibly due to a few bears visiting multiple traps on a single occasion, and second, correlation between individual-specific capture probability, p_i, and individual-specific hair deposition parameter λ_i (see Discussion for further elaboration). To determine the relative influence of each, the hair sample portion of the data were replaced in two scenarios, while maintaining the original capture history structure. To assess the impact of correlation between p_i and λ_i, we first fit a zero-truncated negative binomial distribution to the observed distribution of deposited hair samples, R_{ij}, and then simulated new values for the number of hair samples left per individual/occasion from this distribution (zero-truncated negative binomial parameters were estimated to be size=0.782 and μ=0.857). This removed the potential correlation between p_i and λ_i, but retained the overdispersion. Second, we fit a zero-truncated Poisson distribution to the observed data and then simulated the number of hair samples left per individual/occasion from this distribution (zero-truncated Poisson parameter was estimated to be λ=1.534). This removed both the potential correlation between p_i and λ_i and the overdispersion. As before, 100 data sets were simulated and M_b and M_{b3} were fit to each, but we limited this analysis to the largest missing data scenario (δ=0.25). Time effects for the capture probabilities were detected in the original analysis of this data set (Tredick *et al.*, 2007), but were ruled out as a potential source of bias in the presence of missing data as simulations (not presented here) showed that unmodeled time effects do not interact with missing data to increase bias.

Removing the potential correlation between p_i and λ_i by simulating R_{ij} as negative binomial produced an M_b estimate with slightly less bias relative to our best estimate, but HPD interval coverage of our best estimate when δ=0.25 was still poor (0.16). Fitting M_{b3} to the same scenario decreased bias from -20% to -13% and improved coverage from 0.47 to 0.68. Improved coverage in these scenarios was driven largely by the moderate reduction in bias. Removing the overdispersion by simulating R_{ij} as Poisson did not lead to substantially improved estimates, with mean $\hat{N}$=75 in both cases and similar coverage. In the Discussion we consider other possible sources of bias that may be preventing our methodology from correcting all of the bias in this data set.

Estimates of the behavioral effects also behaved differently then observed in the simulation study. In Simulation 1, the behavioral response $|p - c|$ was increasingly underestimated using M_b

when the level of missing data increased, but estimated with essentially no bias by M_{b3} for the levels of missing data considered. The example data set did not respond to missing data in the same manner. The behavioral response was overestimated with increasing levels of missing data using both M_b and M_{b3}. Removing the potential correlation between p_i and λ_i removed this discrepancy for M_b and reduced, but did not entirely remove, this discrepancy for M_{b3}.

Table 4: Population size estimates, bias (relative to the best estimate of 86), 95% CI coverage, mean 95% CI width, and mean behavioral response estimate when fitting M_b and M_{b3} to the example data set when δ=0.25. In two scenarios, the empirical distribution of the number of hair samples, R, is replaced to remove additional sources of bias.

	M_b					M_{b3}				
R dist.	Mean $\hat{N}$	% Bias	CI Cov.	CI Width	Mean $\lvert\hat{p}-\hat{c}\rvert$	Mean $\hat{N}$	% Bias	CI Cov.	CI Width	Mean $\lvert\hat{p}-\hat{c}\rvert$
Empirical	52	-40	0.02	13.81	0.16	69	-20	0.47	25.52	0.31
Nbinom	57	-34	0.16	21.20	0.13	75	-13	0.68	27.12	0.24
Poisson	.	.	.	.	.	75	-13	0.63	26.15	0.24

REFERENCES

Belant, J.L., Seamans, T.W. & Paetkau, D. (2007) Genetic tagging free-ranging white-tailed deer using hair snares. *The Ohio Journal of Science*, **104**, 50–56.

Best, D., Rayner, J. & Thas, O. (2007) Goodness of fit for the zero-truncated poisson distribution. *Journal of Statistical Computation and Simulation*, **77**, 585–591.

Boulanger, J., McLellan, B.N., Woods, J.G., Proctor, M.F. & Strobeck, C. (2004) Sampling design and bias in DNA-based capture-mark-recapture population and density estimates of grizzly bears. *Journal of Wildlife Management*, **68**, 457–469.

Boulanger, J., Proctor, M., Himmer, S., Stenhouse, G., Paetkau, D. & Cranston, J. (2006) An empirical test of DNA mark-recapture sampling strategies for grizzly bears. *Ursus*, **17**, 149–158.

Dreher, B.P., Rosa, G.J., Lukacs, P.M., Scribner, K.T. & Winterstein, S.R. (2009) Subsampling hair samples affects accuracy and precision of DNA-based population estimates. *The Journal of Wildlife Management*, **73**, 1184–1188.

Frary, V.J., Duchamp, J., Maehr, D.S. & Larkin, J.L. (2011) Density and distribution of a colonizing front of the American black bear Ursus americanus. *Wildlife Biology*, **17**, 404–416.

Henry, P. & Russello, M.A. (2011) Obtaining high-quality DNA from elusive small mammals using low-tech hair snares. *European Journal of Wildlife Research*, **57**, 429–435.

Hwang, W.H. & Huggins, R. (2011) A semiparametric model for a functional behavioural response to capture in capture–recapture experiments. *Australian & New Zealand Journal of Statistics*, **53**, 403–421.

Kendall, K.C. & McKelvey, K.S. (2008) Hair collection. *Noninvasive survey methods for carnivores Island Press, Washington, DC*, pp. 141–182.

Kendall, W.L. (1999) Robustness of closed capture-recapture methods to violations of the closure assumption. *Ecology*, **80**, 2517–2525.

Laufenberg, J.S., Van Manen, F.T. & Clark, J.D. (2013) Effects of sampling conditions on DNA-based estimates of American black bear abundance. *The Journal of Wildlife Management*, **77**, 1010–1020.

Link, W.A. (2013) A cautionary note on the discrete uniform prior for the binomial N. *Ecology*, **94**, 2173–2179.

Link, W.A., Yoshizaki, J., Bailey, L.L. & Pollock, K.H. (2010) Uncovering a latent multinomial: analysis of mark–recapture data with misidentification. *Biometrics*, **66**, 178–185.

Mills, L.S., Citta, J.J., Lair, K.P., Schwartz, M.K. & Tallmon, D.A. (2000) Estimating animal abundance using noninvasive DNA sampling: promise and pitfalls. *Ecological Applications*, **10**, 283–294.

Otis, D.L., Burnham, K.P., White, G.C. & Anderson, D.R. (1978) *Statistical inference from capture data on closed animal populations*, volume 62. Wildlife Society.

Paetkau, D. (2003) An empirical exploration of data quality in dna-based population inventories. *Molecular ecology*, **12**, 1375–1387.

Ramsey, F.L. & Severns, P.M. (2010) Persistence models for mark-recapture. *Environmental and Ecological Statistics*, **17**, 97–109.

Roon, D.A., Waits, L.P. & Kendall, K.C. (2005) A simulation test of the effectiveness of several methods for error-checking non-invasive genetic data. *Animal Conservation*, **8**, 203–215.

Royle, J.A. & Dorazio, R.M. (2008) *Hierarchical modeling and inference in ecology.* Academic Press, London, UK.

Royle, J.A., Karanth, K.U., Gopalaswamy, A.M. & Kumar, N.S. (2009) Bayesian inference in camera trapping studies for a class of spatial capture-recapture models. *Ecology*, **90**, 3233–3244.

Royle, J.A., Magoun, A.J., Gardner, B., Valkenburg, P. & Lowell, R.E. (2011) Density estimation in a wolverine population using spatial capture–recapture models. *The Journal of Wildlife Management*, **75**, 604–611.

Settlage, K.E., van Manen, F.T., Clark, J.D. & King, T.L. (2008) Challenges of DNA-based mark-recapture studies of American black bears. *The Journal of Wildlife Management*, **72**, 1035–1042.

Taberlet, P. (1996) The use of mitochondrial DNA control region sequencing in conservation genetics. *Molecular genetic approaches in conservation*, pp. 125–142. Oxford University Press, New York, NY.

Tanner, M.A. & Wong, W.H. (1987) The calculation of posterior distributions by data augmentation. *Journal of the American statistical Association*, **82**, 528–540.

Tredick, C.A. & Vaughan, M.R. (2009) Dna-based population demographics of black bears in coastal north carolina and virginia. *The Journal of Wildlife Management*, **73**, 1031–1039.

Tredick, C.A., Vaughan, M.R., Stauffer, D.F., Simek, S.L. & Eason, T. (2007) Sub-sampling genetic data to estimate black bear population size: a case study. *Ursus*, **18**, 179–188.

Van Manen, F.T., Mccollister, M.F., Nicholson, J.M., Thompson, L.M., Kindall, J.L. & Jones, M.D. (2012) Short-term impacts of a 4-lane highway on American black bears in eastern North Carolina. *Wildlife Monographs*, **181**, 1–35.

Waits, L.P. & Paetkau, D. (2005) Noninvasive genetic sampling tools for wildlife biologists: a review of applications and recommendations for accurate data collection. *Journal of Wildlife Management*, **69**, 1419–1433.

Wegan, M.T., Curtis, P.D., Rainbolt, R.E. & Gardner, B. (2012) Temporal sampling frame selection in DNA-based capture-mark-recapture investigations. *Ursus*, **23**, 42–51.

White, G.C. (1982) *Capture-recapture and removal methods for sampling closed populations.* Los Alamos National Laboratory, Los Alamos, NM.

Yang, H.C. & Chao, A. (2005) Modeling animals' behavioral response by Markov chain models for capture–recapture experiments. *Biometrics*, **61**, 1010–1017.